Title
ChromTime: modeling spatio-temporal dynamics of chromatin marks.

Journal
Genome biology, 19(1)

ISSN
1474-7596

Authors
Fiziev, Petko
Ernst, Jason

Publication Date
2018-08-01

DOI
10.1186/s13059-018-1485-2

Peer reviewed

Fiziev and Ernst *Genome Biology* (2018) 19:109
https://doi.org/10.1186/s13059-018-1485-2

Genome Biology

METHOD Open Access

ChromTime: modeling spatio-temporal dynamics of chromatin marks

Petko Fiziev[1,2,3] and Jason Ernst[1,2,3,4,5,6*]

Abstract

To model spatial changes of chromatin mark peaks over time we develop and apply ChromTime, a computational method that predicts peaks to be either expanding, contracting, or holding steady between time points. Predicted expanding and contracting peaks can mark regulatory regions associated with transcription factor binding and gene expression changes. Spatial dynamics of peaks provide information about gene expression changes beyond localized signal density changes. ChromTime detects asymmetric expansions and contractions, which for some marks associate with the direction of transcription. ChromTime facilitates the analysis of time course chromatin data in a range of biological systems.

Keywords: Epigenomics, Time course, Spatial dynamics, Histone modifications, Chromatin marks

Background

Genome-wide mapping of histone modifications (HMs) and related chromatin marks using chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) and DNA accessibility through assays for DNase I hypersensitivity (DNase-seq) or transposase-accessible chromatin (ATAC-seq) assays have emerged as a powerful approach to annotate genomes and study cell states [1–5]. Through the efforts of large consortia projects such as ENCODE [6], Roadmap Epigenomics [7], and BLUEPRINT [8] as well as individual labs [9–11], multiple different chromatin marks have been mapped across more than a hundred different cell and tissue types. These maps have yielded numerous insights into gene regulation and genetic and epigenetic association with disease [12–16].

While many mapping efforts have largely focused on single or unrelated cell and tissue types [3, 6], a growing number of biological processes have been studied with temporal epigenomic data using assays such as ChIP-seq, ATAC-seq, or DNase-seq over a time course, which map chromatin marks at consecutive stages during the particular biological process. Such datasets have been generated for a wide range of biological settings, including T-cell development [17], adipogenesis [18], hematopoiesis [19, 20], macrophage differentiation [21], neural differentiation [12], cardiac development [22, 23], somatic cell reprogramming [24–27], embryogenesis [28], and many others [7, 29–37]. The output of these experiments presents a unique opportunity to study the spatio-temporal changes of epigenetic peaks and associated regulatory elements. However, almost all computational methods designed or applied to epigenomic data have been developed based on single or multiple unrelated samples. For example, continuous regions of enrichments of single marks are detected by peak or domain calling methods [38–42]. In cases when multiple chromatin marks are mapped in the same cell type, methods such as ChromHMM [43] and Segway [44] can be used to produce genome-wide chromatin state annotations. In addition, methods have been developed for pairwise comparisons of ChIP-seq signal data by differential peak calling [45, 46].

In the context of time course chromatin data, only a few methods have been proposed that consider temporal dependencies between samples. One such method, TreeHMM [47], produces a chromatin state genome annotation similar to ChromHMM and Segway, while taking into account a tree-like structure that captures lineage relationships between the input cell types in order to potentially derive a more consistent annotation across samples. Another method, GATE [30], produces a

* Correspondence: jason.ernst@ucla.edu
[1]Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA, USA
[2]Department of Biological Chemistry, University of California, Los Angeles, CA, USA
Full list of author information is available at the end of the article

genome annotation based on clustering fixed-length genomic loci that can be modeled with the same switch from one chromatin state to another over time.

One important limitation of methods for pairwise comparison or time course modeling of chromatin data is that they do not directly consider or model spatial changes in the genomic territory occupied by chromatin marks over time. Spatial properties of genomic peaks continuously marked by HMs have gained increasing attention as a potentially important characteristic of chromatin marks. For example, long peaks of H3K27ac have been associated with active cell type-specific locus control regions termed super-enhancers or stretch enhancers in a number of cell types [48, 49]. Also, the length of H3K4me3 peaks has been associated with transcriptional elongation and consistency of cell identity genes [50]. In the context of cancer, long H3K4me3 peaks have been linked to transcriptional elongation and enhancer activity at tumor suppressor genes and have been observed to be significantly shortened in tumor cells [51]. Long H3K4me3 domains have been implicated to mark loci involved in psychiatric disorders [52]. Expanded domains of H3K27me3 and H3K9me3 marks have been shown to be characteristic of terminally differentiated cells compared to stem cells [53]. These studies suggest that length of epigenetic peaks is a dynamic feature that can correlate with activity of putative functional elements regulating specific genes. Computational methods that do not explicitly reason about the spatial changes of chromatin marks have significant limitations for studying the dynamics of these properties because they are unable to detect some territorial changes that might be associated with redistribution of signal or identify asymmetric directional peak boundary movements.

In this work, we present ChromTime, a novel computational method for detection of expanding, contracting, and steady peaks, which can detect patterns of changes in the genomic territory occupied by chromatin mark peaks from time course sequencing data (Fig. 1a). We applied ChromTime to a diverse set of data from different developmental, differentiation, and reprogramming time courses (Table 1). Predicted expansions and contractions in general mark regulatory regions associated with changes in transcription factor (TF) binding or gene expression. ChromTime enables studying the directionality of spatial dynamics of chromatin mark peaks relative to other genomic features, which existing computational approaches do not directly address. Our results show that the direction of predicted expansions and contractions correlates with direction of transcription near transcription start sites (TSSs). ChromTime is a general method that can be used to analyze time course chromatin data from high-throughput sequencing assays such as from ChIP-seq, ATAC-seq, and DNase-seq for a wide range of biological systems to gain insights into the dynamics of gene regulation.

Results

Model for detecting expanding, contracting, and steady peaks from temporal chromatin data

We developed a computational method, ChromTime (https://github.com/ernstlab/ChromTime), designed for systematic detection of expansions, contractions, and steady peaks from time course chromatin data of a single chromatin mark ("Methods"; Fig. 1b). ChromTime takes as input a set of genomic coordinates of aligned sequencing reads from foreground experiments for a chromatin mark and, optionally, control experiments over the time course. The foreground experiments are data from a chromatin sequencing assay such as ChIP-seq, ATAC-seq, or DNase-seq performed at a series of time points. The method consists of two stages—block finding and dynamics prediction. During the block finding stage, ChromTime determines continuous genomic regions (blocks) that may contain peaks of foreground signal enrichment during the time course (Additional file 1: Figure S1A, B). To achieve this, ChromTime partitions the genome into fixed length bins and counts the number of foreground and control reads that map to each bin at each time point. Nearby bins that show significant enrichment are joined into continuous intervals, which subsequently are grouped into blocks if they overlap across time points. As a result, large portions of the genome that are likely to contain background noise at all time points are filtered out, so that peak boundary dynamics are determined within a subset of the genome potentially enriched for the chromatin mark.

During the dynamics prediction stage, for each block, ChromTime determines the most likely positions of the peak boundaries at each time point and whether the peak expands, contracts, or holds steady at each boundary between consecutive time points. The method uses a probabilistic mixture model to partition the signal within each block at each time point into background and peak components (Fig. 1c, Additional file 1: Figure S1C) by reasoning jointly about the data from all time points in the time course. The method assumes that central positions in blocks are more likely to be enriched for foreground reads and thus the peak component is flanked by the background components (Additional file 1: Figure S1D). The number of sequencing reads in bins from each component at each time point is modeled with different negative binomial distributions that can account for the local abundance of control reads. Furthermore, between any two consecutive time points the boundaries of the peaks are assumed to follow one of three possible dynamics: steady, expand, or contract. For steady dynamics, the peak

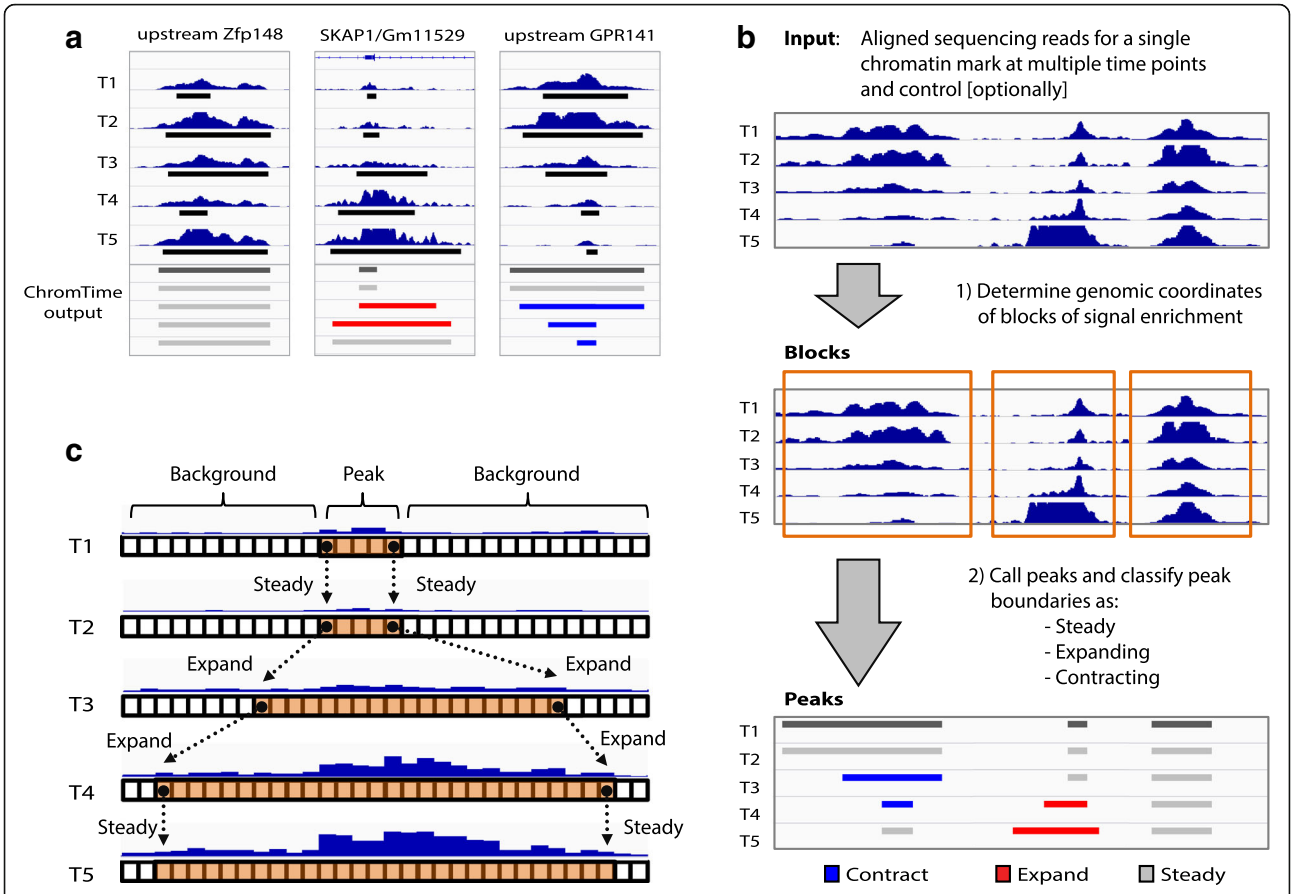

Fig. 1 Overview of the ChromTime method. **a** Examples of H3K4me2 peaks with steady, expanding, and contracting boundary dynamics, shown from left to right, respectively, across five time points during mouse T-cell development [17]. Time points 1, 2, and 3 correspond to in vitro differentiated T-cell precursors (FLDN1, FLDN2a, and FLDN2b), whereas time points 4 and 5 correspond to in vivo purified thymocytes (ThyDN3 and ThyDP). Normalized ChIP-seq signal, MACS2 [38] peaks (*black rectangles*), and ChromTime output are shown for each time point. Peaks upstream of the Zfp148 gene are called steady by ChromTime despite fluctuations of MACS2 peak boundaries. In contrast, ChromTime calls a peak at the Skap1/GM11529 promoter to expand after time points 2 and 3. Conversely, ChromTime calls a peak upstream of the GPR141 gene to contract after time points 2, 3, and 4. **b** Overview of the ChromTime method. During the block-finding stage, input foreground and, optionally, control reads are used to determine blocks of signal enrichment. In the dynamics prediction stage, for each block, peak boundary positions are predicted at each time point and peak boundary dynamics are predicted at each pair of consecutive time points. **c** Predicting dynamics for one block. *Boxes* represent genomic bins at each time point. Foreground signal is depicted as *blue bars* for each bin whose height represents the number of reads mapped to the bin. ChromTime learns a probabilistic mixture model from the input data to partition each block at each time point into peak and background components. Bins in the peak component (*orange*) mark peaks of signal enrichment whereas those in the background component (*white*) mark flanking background signal. The movement of the boundaries on the left and the right side of peaks between consecutive time points is estimated by reasoning jointly about the input data from all time points

boundaries are enforced to have the same genomic position. For expanding and contracting dynamics, the number of genomic bins that the peak boundaries move between the two time points is modeled with different negative binomial distributions which depend on the pair of time points and the corresponding dynamic. ChromTime models time points that have no bins in the peak component with zero length peaks. Thus, appearances of peaks, except at the first time point, are modeled as expansions from zero length peaks and the disappearances of peaks are modeled as contractions to zero length peaks. Each dynamic is also assumed to have a prior probability which captures information about its genome-wide frequency at each time point.

All model parameters are learned jointly from the whole time course. As a result, ChromTime can adapt to different boundary movements, dynamics frequencies, and noise levels across experiments and biological systems. The estimated parameters are used to make a prediction for each block for the most likely positions of the peak boundaries and the corresponding boundary dynamics that had generated the signal within the block. The final output contains predicted peak boundaries annotated and colored by their assigned dynamics, which

Table 1 Datasets used for analysis with ChromTime

System	Chromatin marks	Species	Number of time points	Reference
Adipogenesis	H3K4me2	Mouse	4	[18]
	H3K4me3	Human		
	H3K27ac			
	H3K4me1			
	H3K36me3			
	H3K27me3			
Blood formation	H3K4me2	Mouse	5–7	[19]
	H3K4me3			
	H3K27ac			
	H3K4me1			
	ATAC-seq			
Blood formation	ATAC-seq	Human	5	[20]
Fetal brain development	DNase-seq	Human	3	[7]
Cardiac development	H3K4me3	Mouse	4	[23]
	H3K27ac			
	H3K4me1			
	H3K27me3			
	H3K36me3			
	Pol2			
Cardiac development	H3K4me3	Human	5	[22]
	H3K27me3			
	H3K36me3			
Embryogenesis	H3K4me3	Zebrafish	4	[28]
	H3K27ac			
	H3K4me1			
Macrophage differentiation	H3K4me3	Mouse	5	[21]
	H3K9ac			
	H3K27ac			
	H3K27me3			
Neural differentiation	H3K4me3	Human	5	[12]
	H3K27ac			
	H3K27me3			
	H3K4me1			
Stem cell reprogramming	H3K4me2	Human	4–6	[24]
	H3K4me3			
	H3K27ac			
	H3K4me1			
	H3K27me3			
	H3K36me3			
Stem cell reprogramming	H3K4me2	Mouse	4	[27]
	H3K4me3			
	H3K9ac			
	H3K27ac			
	H3K27me3			
	H3K36me3			
	H3K4me1			
	H3K79me2			
	H3K9me3			
	ATAC-seq			
Stem cell reprogramming	H3K4me3	Mouse	9	[25]
	H3K27ac			
	H3K4me1			
	H3K27me3			
	Pol2			
T-cell development	H3K4me2	Mouse	5	[17]
	H3K(9,14)ac			
	H3K27me3			

can be used for downstream analysis with existing tools and visualized in genome browsers (Fig. 2, Additional file 1: Figure S2; https://github.com/ernstlab/ChromTime).

Reproducibility of ChromTime predictions and association with changes in gene expression, TF binding, and DNaseI hypersensitivity sites

To investigate the reproducibility of ChromTime predictions, we applied ChromTime separately to two biological replicate datasets for the H3K4me2 and H3K(9/14)ac marks in T-cell development in mouse [17] and confirmed, on average, strong enrichment for the same ChromTime annotations co-localizing across replicates (Additional file 1: Figure S3). We then applied the method to data from pooled replicates for the H3K4me2 mark from the mouse T-cell development study [17], to data for the H3K4me3 and H3K27ac marks from a study on stem cell reprogramming in human [24], to ATAC-seq data from a mouse stem cell reprogramming time course [27], and to a human fetal brain development time course that we constructed from DNase-seq datasets from Roadmap Epigenomics [7]. To investigate the biological relevance of ChromTime predictions, for blocks overlapping TSSs we examined changes in the corresponding gene expression. Peaks with predicted expanding and contracting boundaries that overlap annotated TSSs associated with increases and decreases, respectively, in gene expression (Fig. 3, Additional file 1: Figure S4). Additionally, for all chromatin marks we examined enrichments of TF binding sites across all blocks [6, 17, 27], and in the case of HMs, also enrichments of DNaseI hypersensitivity sites (DHSs) [7]. Predicted peaks with expanding and contracting boundaries

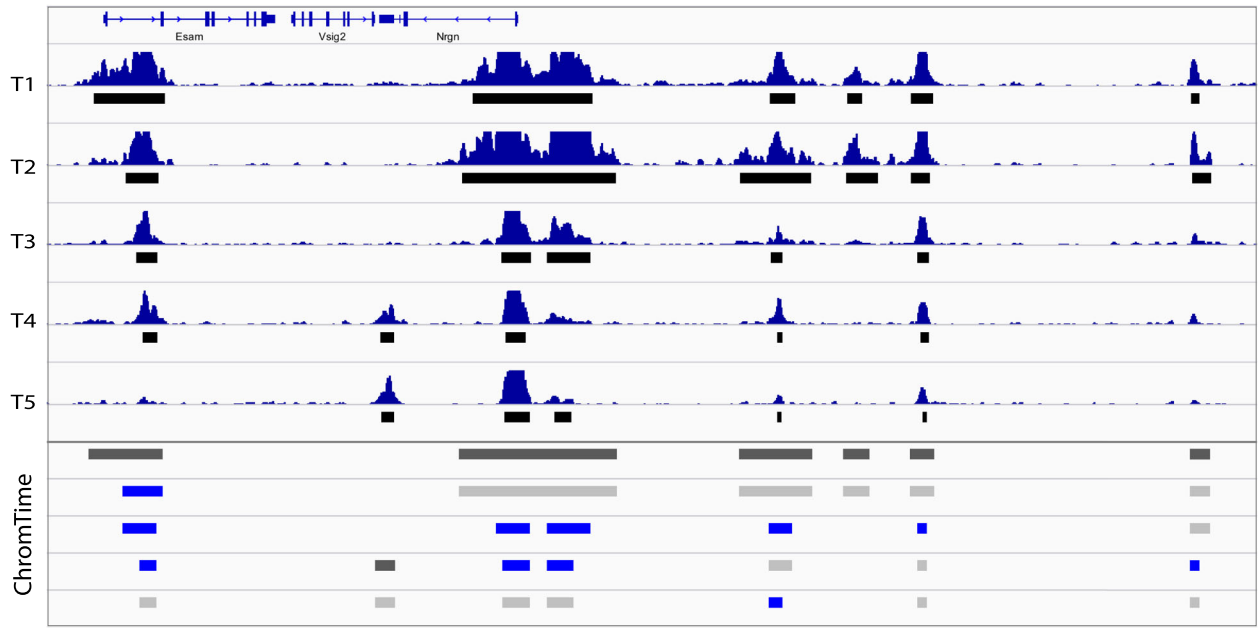

Fig. 2 Sample output from ChromTime with contracting peaks. Genome browser screenshot with sample output of ChromTime for H3K4me2 from the T-cell development time course in mouse [17] with five time points at the Esam/Vsig2/Nrgn locus. Time points 1, 2, and 3 correspond to in vitro differentiated T-cell precursors (FLDN1, FLDN2a, and FLDN2b), whereas time points 4 and 5 correspond to in vivo purified thymocytes (ThyDN3 and ThyDP). The input ChIP-seq signal and MACS2 [38] peaks (*black boxes* under each signal track) are shown in the *upper panel* of the screenshot. The ChromTime-predicted peaks colored by their boundary dynamics for each block at each time point are shown in the *bottom panel*. The first peak in each block is colored in *dark gray*. Each subsequent peak is colored with respect to the predicted dynamic relative to its previous time point. Peaks with steady boundaries on both sides are shown in *light gray*, and those with at least one contracting boundary are shown in *blue*. Nearby peaks that touch boundaries are visualized as one peak by the genome browser. Not shown in the figure are expanding peaks, peaks at single time points, and peaks with opposite dynamics (EXPAND on the *left* and CONTRACT on the *right*, or vice versa), which would be colored in *red*, *orange*, and *black*, respectively. See Additional file 1: Figure S2 for examples of predicted expanding peaks

were enriched for sites bound by important transcriptional regulators in each biological system in a cell type-specific manner. Expanding and contracting HM peaks were also enriched for cell type-specific DHSs. Furthermore, peaks with predicted steady boundaries showed enrichment for TF binding sites that are shared between the first and the last time point in the corresponding time courses, which mark potentially stable regulatory elements. Similar enrichments in the case of HM peaks were also seen for shared DHSs.

Predicted spatial dynamics by ChromTime associate better with gene expression changes compared to boundary position changes of peaks called from individual time points in isolation

We next investigated whether ChromTime's approach for reasoning jointly about the whole time course increases power to detect associations with gene expression compared to considering boundary differences of peaks at consecutive time points called in isolation. Specifically, we analyzed gene expression changes of genes with TSSs overlapping ChromTime peaks in relation to posterior probabilities for expansions and contractions compared to boundary differences of peaks called with ChromTime from data from individual time points in isolation. We investigated this in the context of H3K4me2 peaks in mouse T-cell development [17] and for H3K4me3 peaks in stem cell reprogramming in human [24]. In most cases, ranking boundary changes of peaks in blocks with at least one non-zero length peak by their predicted ChromTime posterior probabilities for expansions and contractions associated, on average, with larger gene expression changes compared to ranking boundaries directly based on the change in the genomic positions of the boundaries of ChromTime peaks called at individual time points in isolation (Fig. 4, Additional file 1: Figure S5A). These results also held when using peaks from two different peak callers, MACS2 [38] and SICER [40], applied on data from individual time points (Additional file 1: Figure S5B, C).

Spatial dynamics contain information about gene expression changes between consecutive time points not captured by corresponding pairwise signal density changes

We next investigated whether there is additional information in ChromTime predictions with respect to gene

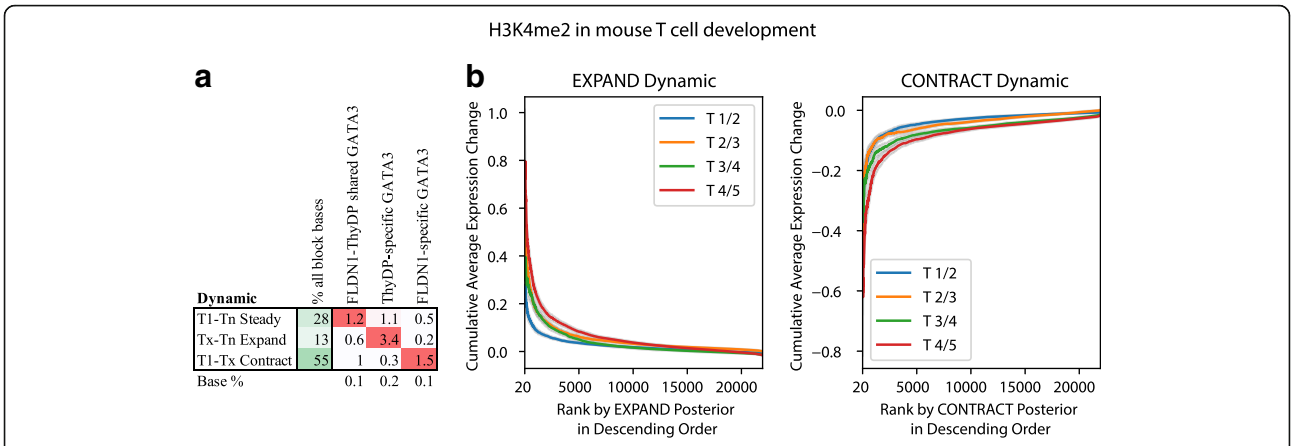

Fig. 3 Changes in GATA3 binding and gene expression at predicted H3K4me2 dynamics in T-cell development. **a** Fold enrichments of cell type-specific and shared peaks of GATA3, which is a master regulator in T-cell development [17], are shown for three sets of blocks with predicted H3K4me2 peaks: 1) blocks with peaks present at all time points whose boundaries hold steady on both sides throughout the whole time course (*T1-Tn Steady*); 2) blocks with non-contracting peaks whose boundaries expand between at least one pair of consecutive time points and have a peak at the last time point (*Tx-Tn Expand*); and 3) blocks with non-expanding peaks whose boundaries contract between at least one pair of consecutive time points and have a peak at the first time point (*T1-Tx Contract*). The first column shows the percentage of bases out of all bases covered by peaks of the set. The last row shows the baseline percentage for each feature out of all bases covered by ChromTime peaks at any time point. Percentages are colored from 0 (*white*) to 100 (*green*). Fold enrichments in each column are colored from 1 (*white*) to the maximum value in the column (*red*). FLDN1 and ThyDP denote differentiated T-cell precursors and purified thymocytes, which are the first and the last time point, respectively. **b** Boundaries of predicted H3K4me2 peaks in blocks with at least one predicted non-zero length peak overlapping annotated TSSs were sorted in decreasing order by their posterior probability for EXPAND dynamic (*left plots*) and CONTRACT dynamic (*right plots*) at each pair of consecutive time points (Additional file 2: Supplementary methods). Gene expression differences between consecutive time points were calculated as the average difference across all genes with overlapping TSSs for each block. For each posterior rank (x-axis) the plot shows the cumulative average gene expression difference across all peak boundaries with equal or higher posterior probabilities (y-axis). Expanding boundaries associated with increase of gene expression and contracting boundaries associated with decrease of gene expression. *Shaded regions* correspond to 95% confidence intervals

expression changes beyond what can be captured by pairwise signal density changes or by differential peak calls. For this analysis, we focused on H3K4me2 in mouse T-cell development [17] and H3K4me3 in human stem cell reprogramming [24]. For pairs of consecutive time points, we computed the change in signal density in the region starting at the left-most and ending at the right-most predicted peak boundary in the block (Additional file 2: Supplementary methods). We associated the signal density changes with gene expression changes at the nearest TSS within 50 kb of each block and computed the average gene expression change as a function of the signal density change within blocks (Fig. 5). We found that locations with the same signal density change can associate with significantly different average gene expression changes of proximal genes depending on the predicted ChromTime dynamics. Notably, bidirectional expansions, expansions occurring on both sides of a peak, associated for a range of signal density changes with greater average increase in gene expression than unidirectional expansions, those expansions occurring on one side but steady on the other, when controlling for the signal density change. These unidirectional expansions in turn associated for a range of signal density changes with greater expression change than steady regions, those regions with a steady call on both sides of a peak, when controlling for the signal density change. We observed a similar relationship for contractions and decrease of gene expression. These results were replicated also after substituting ChIP-seq signal density changes with differential peak scores from two differential peak calling methods, SICER [40] and MACS2 [38] (Additional file 1: Figure S6A, B). Therefore, ChromTime predictions can provide additional information about gene expression changes beyond what is contained in the corresponding signal density changes as measured directly or by utilizing differential peak-calling procedures.

Spatial dynamics are correlated between multiple chromatin marks

Previous studies have shown that the locations of different chromatin marks can be correlated [3, 54]. In this context, we tested whether multiple chromatin marks can also exhibit jointly the same type of spatio-temporal dynamics. For this purpose, we compared the genomic locations of predicted expansions, contractions, and steady peaks for different chromatin marks within the same time course. We focused on three previously published time courses—stem cell reprogramming in human

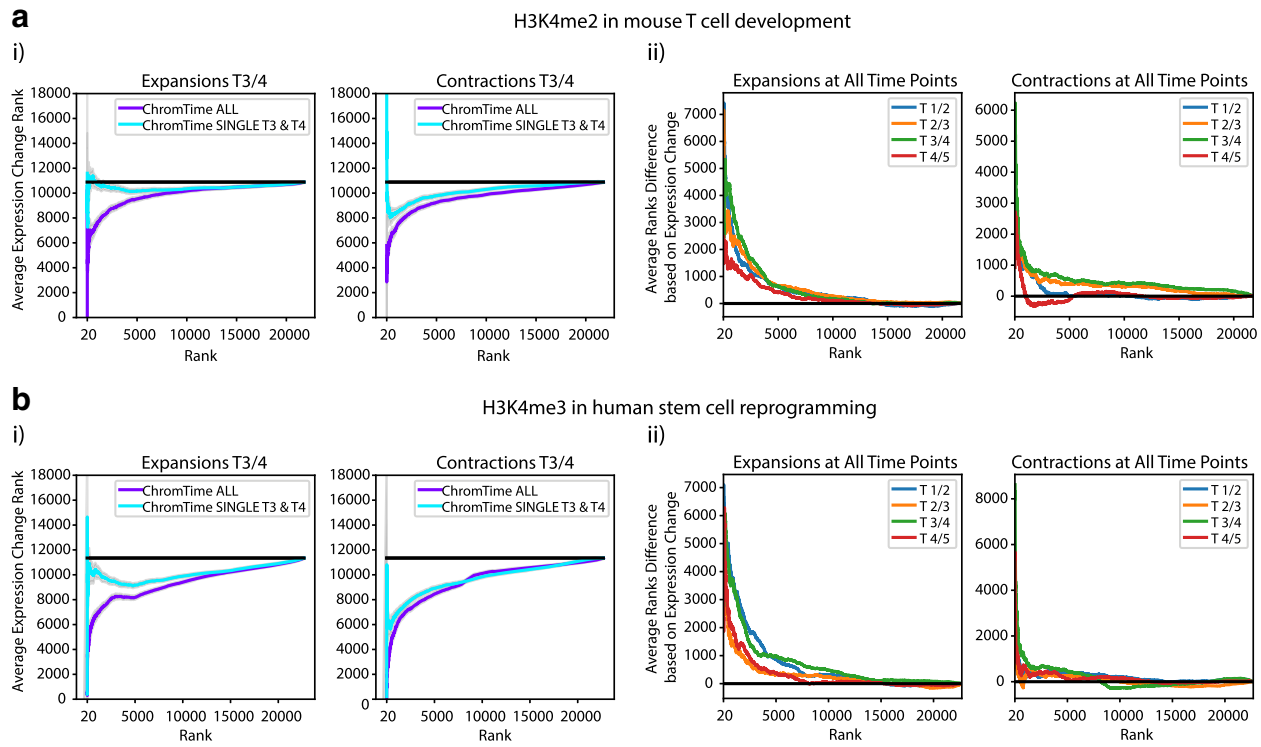

Fig. 4 ChromTime predictions associate better with expression changes than boundary movements of peaks called in isolation. **a** For H3K4me2 in mouse T-cell development [17] ChromTime was applied once with data from all time points (ChromTime ALL), and once with single time points in isolation (ChromTime SINGLE; Additional file 2: Supplementary methods). Time points 1, 2, and 3 correspond to T-cell precursors, 4 and 5 to purified thymocytes. Peaks called by both procedures overlapping annotated TSSs were analyzed for their relationship with gene expression changes. *i Left*: Comparison of agreement with expression for expansions when applying ChromTime ALL and ChromTime SINGLE for the change between time points 3 and 4. Peak boundaries were sorted in decreasing order of their EXPAND posterior probabilities from ChromTime ALL and compared to sorting them in decreasing order of the difference of peak boundary positions in ChromTime SINGLE peaks with positive differences in boundary positions indicating peaks expanding with time. Each boundary was also ranked by the average gene expression difference of TSSs overlapping its block in decreasing order with positive expression differences indicating gain with time. The cumulative average boundary rank of expression change (y-axis) is shown for the boundary change ranking for ChromTime ALL and ChromTime SINGLE (x-axis). Low Y-values indicate stronger association with expression changes. *Black line* shows expected average expression change rank. *Shaded regions* indicate 95% confidence intervals. Plots for other time points can be found in Additional file 1: Figure S5. *Right*: Analogous to left plots for contract posterior probabilities for ChromTime ALL, increasing order of the difference of boundary change positions for ChromTime SINGLE, and increasing order of expression changes. *ii* Differences between ChromTime ALL and ChromTime SINGLE values shown in *i* between time points 3 and 4 as well as for all other pairs of time points. Positive values correspond to boundary ranks for which ChromTime ALL posteriors better associate with gene expression changes than boundary movements of ChromTime SINGLE peaks. *Black lines* show expected difference of zero between random rankings. **b** As in **a** for H3K4me3 in human stem cell reprogramming [24]. Time points correspond to human inducible and immortalized fibroblast-like (hiF-T) cells, hiF-T at 5, 10, and 20 days after induction, and human induced pluripotent stem cells (hIPSC)

[24], stem cell reprogramming in mouse [27], and adipogenesis in mouse [18]—where multiple chromatin marks were mapped (Fig. 6, Additional file 1: Figure S7). In all three datasets, we observed that predicted expansions co-localized preferentially for H3K4me2, H3K4me3, and H3K27ac and to a lesser extent for H3K4me1 and similarly for predicted contractions and steady peaks. In contrast, different predicted spatial dynamics for H3K36me3 and H3K27me3 tended to occupy distinct locations. In addition, in mouse reprogramming [27], ChromTime predicted dynamics of ATAC-seq, H3K4me2, H3K4me3, H3K27ac, H3K9ac, and, to a lesser extent, of H3K4me1 and H3K79me2 peaks co-localized preferentially (Additional file 1: Figure S7). These results suggest that spatial dynamics of chromatin marks are coordinated at least at a subset of genomic locations.

Direction of expansions and contractions is correlated with direction of transcription

ChromTime can predict unidirectional expansions and contractions, which enables analysis of directionality of spatial dynamics of peaks, an aspect of chromatin regulation that has not been previously systematically explored. To investigate this, we applied ChromTime on data from 13 previously published studies from a

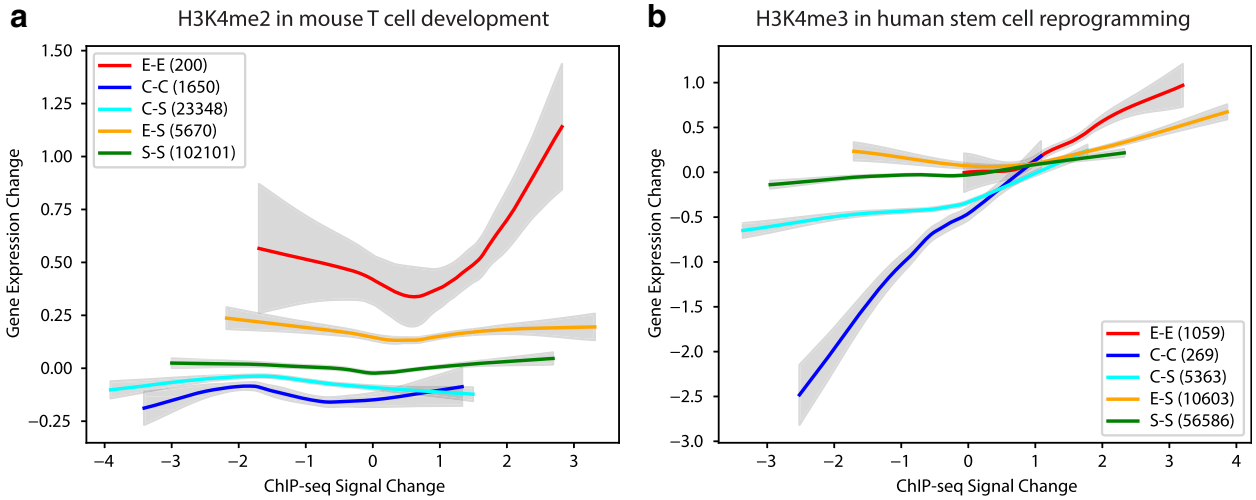

Fig. 5 Spatial dynamics can contain additional information about gene expression changes beyond signal density changes. Gene expression change is plotted as function of ChIP-seq signal density change after loess smoothing for each predicted ChromTime dynamic for **a** H3K4me2 dynamics in T-cell development in mouse [17] and **b** H3K4me3 dynamics in stem cell reprogramming in human [24] (Additional file 2: Supplementary methods). Peaks of each type of dynamics were pooled from all time points in each dataset for this analysis. Peaks with asymmetric dynamics E/S and S/E were pooled together in the "E-S" group. Similarly C/S and S/C peaks were pooled in the "C-S" group. The total number of peaks in each group is shown in parenthesis. In both systems, for a range of signal density changes, peaks with the same signal density change associated with different gene expression changes depending on the predicted spatial dynamic. *Shaded regions* represent 95% confidence intervals

variety of developmental, differentiation, and reprogramming processes (Table 1) for nine different HMs, including narrow and broad marks, and for Pol2, ATAC-seq, and DNase-seq. We observed that unidirectional expansions and contractions are predicted in most cases, on average, to be the majority of all expansions and contractions, respectively, at a given pair of consecutive time points (Additional file 1: Figure S8). One hypothesis for the prevalence of asymmetric boundary movements for the promoter-associated chromatin marks is that the direction of boundary movements is associated with the asymmetry of transcription initiation in promoter regions. To test this hypothesis, for each dataset we compared the prevalence of each class of unidirectional dynamics as a function of its distance to the nearest annotated TSS and the orientation of the corresponding gene (Fig. 7). Consistent with our hypothesis, for H3K4me3, H3K4me2, H3K(9/14)ac, H3K79me2, and Pol2, we found that unidirectional expansions that expand into the gene body (i.e., in the same direction as transcription) were substantially more frequently found in proximity of TSSs compared to unidirectional expansions in the opposite direction. Moreover, this difference was not observed for expansions that are distal from TSSs. Similarly, in most cases for these marks unidirectional contractions that contract towards the TSS of the nearest gene (i.e., in the opposite direction of transcription) were substantially more frequent compared to unidirectional contractions in the opposite direction in proximity of TSSs, whereas their frequencies at distal sites showed much smaller differences. HMs H3K27ac, H3K4me1, and H3K27me3 and ATAC-seq and DNase-seq exhibited the same trend, but to a lesser degree.

Discussion

In this work, we presented ChromTime, a novel computational method for systematic detection of expanding, contracting, and steady peaks of chromatin marks from time course high-throughput sequencing data. ChromTime employs a probabilistic graphical model that directly models changes in the genomic territory occupied by single chromatin marks over time. This approach allowed us to directly encode our modeling assumptions about dependencies between variables in an interpretable and extendable framework.

We applied ChromTime on ChIP-seq data for broad and narrow HMs and for Pol2, and on ATAC-seq and DNase-seq data from a variety of developmental, differentiation, and reprogramming courses. Our results show that the method can identify sets of expanding and contracting peaks that are biologically relevant to the corresponding systems. In particular, expansions and contractions associate with up- and down-regulation of gene expression and differential TF binding, supporting the biological relevance of ChromTime predictions.

ChromTime gains power by both reasoning jointly about all time points in a time course and by explicitly modeling the peak boundary movements. Supporting

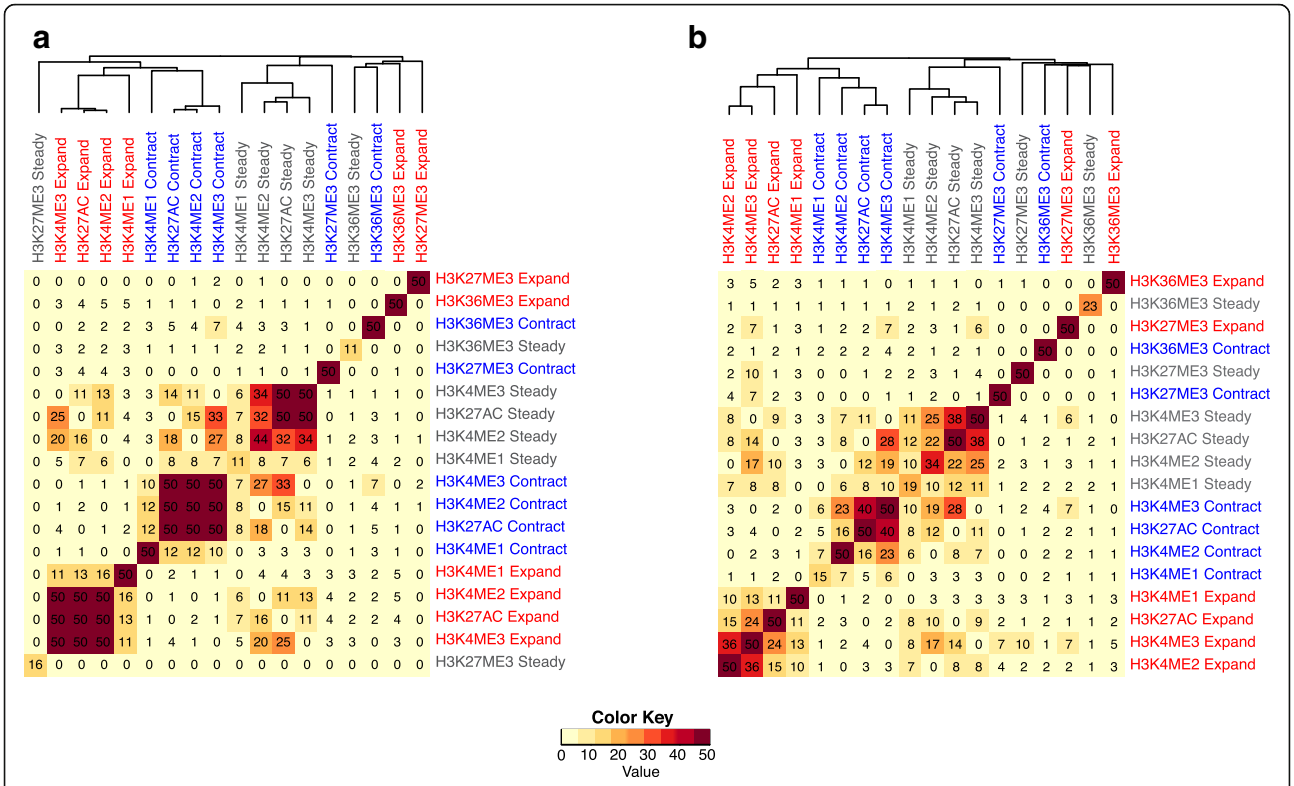

Fig. 6 Spatial dynamics of multiple different chromatin marks co-localize within a time course. Hierarchical clustering with optimal leaf ordering [80] of the geometric average fold enrichments taken across all time points of the overlap of every pair of predicted spatial dynamics for mapped HMs in **a** mouse adipogenesis [18] and **b** human stem cell reprogramming [24]. At each pair of time points, "Expand" and "Contract" dynamics are defined as all peaks that are predicted as either unidirectional or bidirectional expansions and contractions, respectively, whereas "Steady" dynamics are defined as all peaks that have predicted steady boundaries at both sides. Peaks with "Expand" dynamic on one side and "Contract" dynamic on the other were excluded from this analysis. In both datasets, expansions, contractions, and steady peaks of H3K4me2, H3K4me3, and H3K27ac and, to a lesser extent, of H3K4me1 tend to cluster together within each of the three classes, whereas spatial dynamics of H3K27me3 and H3K36me3 peaks tend to occupy different locations. All enrichments were capped at 50 before clustering

this, in our analyses we observed that territorial changes identified by ChromTime had better agreement with gene expression changes compared to considering directly the boundary change of peaks called on data from individual time points in isolation. Additionally, we also observed for a range of cases that expanding and contracting peaks associated, on average, with greater change in gene expression compared to peaks with steady boundaries even after controlling for signal density changes. Some of the power that ChromTime gains from considering spatial information might be explained by its ability to differentiate territorial expansions or contractions, which can reflect changes in the number of TF binding sites in close vicinity, from changes in signal density within steady peak boundaries. Changes in signal density without territorial expansions or contractions might reflect a change in the proportion of cells with the chromatin mark without large changes in activity in any one cell. Additional power can come from the temporal and spatial information that allows the model to effectively smooth over noise in the data, thus enabling more biologically relevant inferences.

ChromTime enables novel analysis of directionality of spatial epigenetic dynamics. In this context, we found that asymmetric unidirectional expansions and contractions for several marks correlate strongly with direction of transcription in promoter proximal regions, which suggests that spatial dynamics at such locations may be related to actions of the transcriptional machinery. One possible explanation for the observed correlation between the direction of spatial dynamics of at least some HMs and transcription can be provided in part by previous studies that have shown that the Pol2 elongation machinery can recruit H3K4-methyltransferases, such as members of the SET [55] and MLL [56] families, at the promoters of genes. Our findings are consistent with such models where the Pol2 complex itself may be facilitating the attachment and removal of these marks [57].

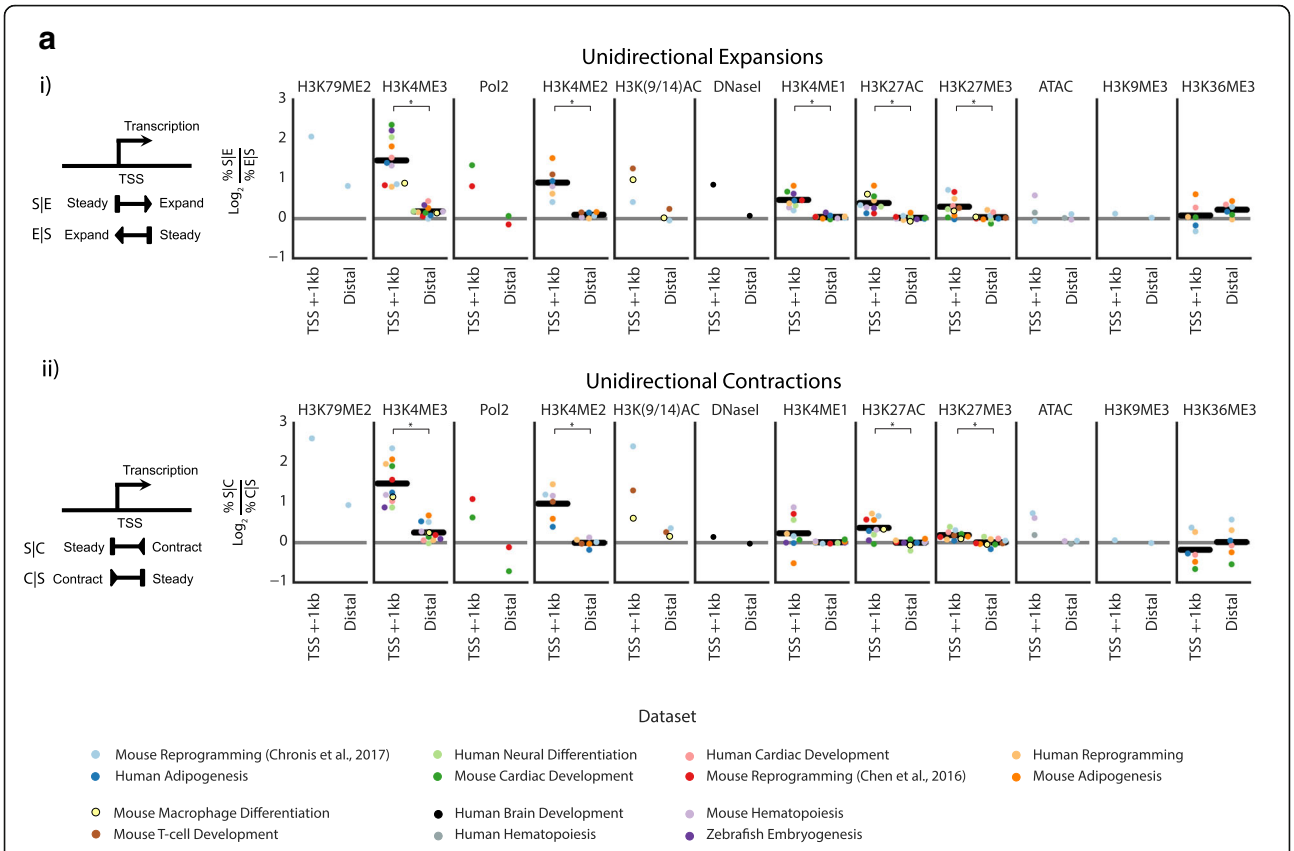

Fig. 7 Direction of asymmetric dynamics correlates with direction of transcription. **a** *i Left panel* shows a schematic representation of unidirectional expansions that expand in the same direction as transcription and in the opposite direction of transcription. The adjacent plots show, for each mark, the average $\log_2$ ratio across all time points in each time course between the fraction of unidirectional expansions that expand in the same directions as transcription of the nearest gene and the fraction of unidirectional expansions that expand in the opposite direction of transcription of the nearest gene, separately for blocks that are within 1 kb of annotated TSSs and for more distal blocks. Positive values correspond to enrichment of unidirectional expansions in the same direction as transcription. For marks mapped in at least six time courses, a *black line* is plotted representing the average across all data sets and significant differences are denoted with *asterisks* based on a two-tailed Mann-Whitney test at a *P* value threshold of 0.05. *ii Left panel* shows analogous schematic for unidirectional contractions. Likewise, adjacent plots show, for each mark, the average $\log_2$ ratio between the fraction of unidirectional contractions that contract in the opposite direction of transcription of the nearest TSS and unidirectional contractions that contract in the same direction as transcription of the nearest TSS. **b** *Left panel* shows an example of unidirectional expansions between pairs of time points that expand in the same direction as transcription at the Hs6st1 gene of the H3K4me2 mark in the T-cell development dataset [17]. Right panel shows an example of unidirectional contractions in the opposite direction of transcription at the DNMT3B gene. Time points 1, 2, and 3 correspond to in vitro differentiated T-cell precursors, whereas time points 4 and 5 correspond to in vivo purified thymocytes. The predicted ChromTime peaks colored by their boundary dynamics are shown under the signal track for each time point

The ChromTime software is also relatively efficient in terms of runtime, particularly when using its option to parallelize all computations during the parameter learning and prediction phases over multiple CPU cores. In our tests, processing ChIP-seq data for the H3K4me2 mark and control data from five time points in mouse T-cell development [17] took 3 h on a laptop computer using four CPU cores.

We applied ChromTime to a range of data types but found no single setting of the method options to be preferable in all cases ("Methods"). We thus created three modes with different default options: punctate mode used for ATAC-seq and DNase-seq, narrow mode used for ChIP-seq of narrow HMs, and broad mode used for ChIP-seq of broad HMs and Pol2. In principle, ChromTime can also be applied on ChIP-seq data of sequence-specific TFs in punctate mode. However, for these data, where binding can often be associated with a single point source such as individual instances of DNA sequence regulatory motifs, methods that predict the single point source across time points and the binding intensity associated with the source at each time point may be a more natural way to model the data.

Another limitation of the ChromTime method is that while the runtime of ChromTime still scales linearly with the number of time points, T, the number of observed combinations of dynamics can scale exponentially with T. This exponential growth can complicate downstream analyses that directly consider each combination of dynamics, as there will be 3^{T-1} possible sequences of dynamics at each side of a peak. Extensions of the ChromTime model could model the large number of combinations as being instances of a smaller number of more distinct dynamic patterns.

Conclusions

The increasing availability of time course chromatin data provides an opportunity to understand chromatin dynamics in many biological systems. To facilitate reaching this goal we developed ChromTime, which systematically detects expanding, contracting, and steady peaks, allowing extraction of additional information from these data. ChromTime gains power by both reasoning about data from all time points in the time course and by explicitly modeling movements of peak boundaries. We showed that ChromTime predictions associate with relevant genomic features such as changes in gene expression and TF binding. We demonstrated that territorial changes of peaks can contain additional information beyond signal density changes with respect to gene expression of proximal genes. ChromTime allows for novel analysis of directionality of spatial dynamics of chromatin marks. In this context, we showed for multiple chromatin marks that the direction of predicted asymmetric expansions and contractions of peaks strongly associates with direction of transcription in proximity of TSSs. ChromTime is generally applicable to modeling time courses of chromatin marks and thus should be a useful tool to gaining insights into dynamics of epigenetic gene regulation in a range of biological systems.

Methods

Overview of the ChromTime method

ChromTime takes as input a set of files in BED format with genomic coordinates of aligned sequencing reads from experiments for a single chromatin mark from a high-throughput sequencing experiment such as ChIP-seq, ATAC-seq, or DNase-seq over a time course and, optionally, from a set of control experiments. ChromTime consists of two stages (Fig. 1b, c):

1. Detecting genomic intervals (blocks) potentially containing regions of signal enrichment (peaks)
2. Learning a probabilistic mixture model for boundary dynamics of peaks within blocks throughout the time course and computing the most likely spatial dynamic and peak boundaries for each block throughout the whole time course

Detecting genomic blocks containing regions of signal enrichment

The aim of this stage is to determine approximately the genomic coordinates of regions with potential peaks of signal enrichment at any time point in the time course (Additional file 1: Figure S1A, B). The signal within these blocks will be used as input to build the mixture model in the next stage of ChromTime. ChromTime supports analysis of punctate, narrow, and broad marks in three different modes, which are defined by different default options. The method partitions the genome into non-overlapping bins of predefined length, BIN_SIZE (by default, 200 bp in narrow and punctate modes, 500 bp in broad mode) and counts for each bin and time point the number of sequencing reads whose alignment starting positions after shifting by a predefined number of bases (SHIFT, 100 bp in the direction of alignment by default) are within its boundaries. Next, each bin at each time point is tested for enrichment based on a Poisson background distribution at a predefined false discovery rate (FDR; 0.05 by default). The expected number of reads for a bin at position p and time point t, $\lambda_{t,p}$, in the Poisson test is computed conservatively as the maximum of:

1) If control reads are provided: for each window of size w = 1000 bp, 5000 bp, and 20,000 bp the average number of control reads in the window

centered at the current bin, normalized by the ratio of total reads in the foreground and control experiments, that is:

$$\lambda_{t,p,w} = \frac{\#[\text{Total Foreground Reads}]}{\#[\text{Total Control Reads}]} \frac{\text{BIN\_SIZE}}{w} \text{Ctrl}_{t,p,w}$$

where $\text{Ctrl}_{t,p,w}$ is the total number of control reads in each window of size w around the bin at position p at time point t.

2) The average number of foreground reads per genomic bin.
3) One read.

Testing multiple different window sizes for the background is a strategy we adopted from the MACS2 peak caller [38].

Within each time point, consecutive bins that are significantly enriched are merged into continuous intervals. The intervals are further extended in both directions to include continuous stretches of bins where each bin is significant based on a Poisson background distribution at a weaker P value threshold (0.15 by default). Extended intervals within a predefined number of non-significant bins, MAX_GAP (3 bins by default), are further joined together. This joining strategy has been previously implemented by other peak callers for single datasets such as SICER [40]. Next, overlapping intervals across time points are grouped into blocks. To capture more of the potential background signal and to increase the likelihood that central bins within blocks contain higher foreground signal, the start and end positions of each block are extended additionally by a predefined number of bins, BLOCK_EXTEND (5 bins by default), upstream of the left-most coordinate and downstream of the right-most coordinate of the intervals in the block, respectively, or up to the middle point between the current block and its adjacent blocks if they are within BLOCK_EXTEND bins apart. Restricting BLOCK_EXTEND to a relatively limited number of bins helps to keep the running time of the method within reasonable bounds.

In narrow and punctate modes, blocks that contain multiple intervals at the same time point separated by gaps of non-significant bins longer than MAX_GAP are split into sub-blocks at each gap between those intervals. In particular, all gaps within a block are intersected across the time points that have gaps. For each gap intersection, the block is split at the position with the lowest average foreground signal across all time points. In broad mode, no such splitting is performed in order to avoid excessive peak fragmentation.

Probabilistic mixture model for boundary dynamics of peaks within blocks across the time course

The foreground and the expected signal within the blocks are used as input to build a probabilistic mixture model for the boundary dynamics of the peaks within blocks (Additional file 1: Figure S1C). One core assumption of the model is that each block contains at each time point exactly one peak, which can potentially have a length of zero bins. This implies that, at each time point, the bins within a block can be partitioned into three continuous intervals: left-flanking background, foreground peak, and right-flanking background. For the bin in block i, at time point t and position p, let $O_{i,t,p}$ denote the random variable that models the number of observed foreground reads, and let $o_{i,t,p}$ denote the corresponding observed read counts. Let $V_{i,t,p}$ denote the random variable for the label of the corresponding bin, which can either have the value PEAK or BACKGROUND. Let $X_{i,t,p}$ denote a random variable for the vector of covariates for the corresponding bin, and $x_{i,t,p}$ their corresponding values. The distribution of $O_{i,t,p}$ conditioned on $V_{i,t,p}$ and $X_{i,t,p}$ is modeled with different negative binomial distributions depending on the value of $V_{i,t,p}$ and $x_{i,t,p}$:

$$\begin{aligned} P(O_{i,t,p} = o_{i,t,p} | V_{i,t,p} = \text{PEAK}, X_{i,t,p} = x_{i,t,p}) &= \text{NB}(o_{i,t,p}; \mu_{\text{PEAK},i,t,p}, \delta_t) \\ &= \frac{\Gamma(o_{i,t,p} + \delta_t)}{o_{i,t,p}!\Gamma(\delta_t)} \times \left(\frac{\delta_t}{\mu_{\text{PEAK},i,t,p} + \delta_t}\right)^{\delta_t} \\ &\quad \times \left(\frac{\mu_{\text{PEAK},i,t,p}}{\mu_{\text{PEAK},i,t,p} + \delta_t}\right)^{o_{i,t,p}} \end{aligned}$$

and

$$\begin{aligned} P(O_{i,t,p} = o_{i,t,p} | V_{i,t,p} = \text{BACKGROUND}, X_{i,t,p} = x_{i,t,p}) &= \text{NB}(o_{i,t,p}; \mu_{\text{BACKGROUND},i,t,p}, \delta_t) \\ &= \frac{\Gamma(o_{i,t,p} + \delta_t)}{o_{i,t,p}!\Gamma(\delta_t)} \times \left(\frac{\delta_t}{\mu_{\text{BACKGROUND},i,t,p} + \delta_t}\right)^{\delta_t} \\ &\quad \times \left(\frac{\mu_{\text{BACKGROUND},i,t,p}}{\mu_{\text{BACKGROUND},i,t,p} + \delta_t}\right)^{o_{i,t,p}} \end{aligned}$$

where δ_t is the dispersion parameter. Similarly to negative binomial regression models [58], ChromTime models the mean of each component through the log link as a linear combination of a two-dimensional vector of covariates, $x_{i,t,p} = (1, \log \lambda_{i,t,p})$, which includes a constant term and the logarithm of the expected number of reads in the bin as computed in the previous section:

$$\mu_{\text{PEAK},i,t,p} = \exp\left[\alpha_t + \gamma_t \log \lambda_{i,t,p}\right]$$

$$\mu_{\text{BACKGROUND},i,t,p} = \exp\left[\beta_t + \gamma_t \log\lambda_{i,t,p}\right]$$

where α_t, β_t and γ_t are time point-specific scalar parameters. Negative binomial distributions have been successfully employed in a similar manner to capture the over-dispersion of sequencing reads in peak callers for single samples such as ZINBA [59]. Of note, however, ChromTime requires that the dispersion parameter δ_t and the coefficient γ_t are shared between the two components at each time point. The first requirement ensures that the distribution with the smaller mean value has higher probabilities compared to the distribution with the larger mean value for the lowest values of the support domain of the negative binomial distribution, and that the opposite holds for the largest values of the support domain (Additional file 2: Supplementary methods). Sharing the dispersion parameter here is analogous to sharing the variance parameter in Gaussian mixture models. The second requirement to share the γ_t parameter ensures that the control signal has equal importance in each component.

Formally, let $B_{i,L,t}$ and $B_{i,R,t}$ denote the random variables corresponding to the first and the last bin, respectively, in the peak partition at time t for block i relative to the beginning of the block, and let N_i be the length of the block. We then have $1 \le B_{i,L,t} \le N_i + 1$ and $0 \le B_{i,R,t} \le N_i$, with values of $B_{i,L,t} = N_i + 1$ and $B_{i,R,t} = 0$ corresponding to the special cases of starting a peak after all positions and ending a peak before all positions in a block, respectively. For $B_{i,L,t}$ and $B_{i,R,t}$ to denote valid interval boundaries, ChromTime also requires that $B_{i,L,t} \le B_{i,R,t} + 1$ at each time point. These constraints can be formally encoded by introducing one auxiliary binary variable for each time point in the model, $Z_{i,t}$, such that:

$$P\left(Z_{i,t} = 1 | B_{i,L,t} = l, B_{i,R,t} = r\right) = \begin{cases} 1 & \text{if } 1 \le l \le r + 1 \le N_i + 1 \\ 0 & \text{otherwise} \end{cases}$$

and thus also

$$P\left(Z_{i,t} = 0 | B_{i,L,t} = l, B_{i,R,t} = r\right) = \begin{cases} 0 & \text{if } 1 \le l \le r + 1 \le N_i + 1 \\ 1 & \text{otherwise} \end{cases}$$

ChromTime treats all $Z_{i,t}$ variables as observed with values equal to 1 for all blocks and time points.

The conditional probability of the bin labels, $V_{i,t,p}$, given the peak boundaries, $B_{i,L,t}$ and $B_{i,R,t}$, are defined to be:

$$P\left(V_{i,t,p} = \text{PEAK} | B_{i,L,t} = l, B_{i,R,t} = r\right) = \begin{cases} 1 & \text{if } l \le p \le r \\ 0 & \text{otherwise} \end{cases}$$

and thus also

$$P\left(V_{i,t,p} = \text{BACKGROUND} | B_{i,L,t} = l, B_{i,R,t} = r\right) = \begin{cases} 0 & \text{if } l \le p \le r \\ 1 & \text{otherwise} \end{cases}$$

The probability of the observed read counts at time t, $\boldsymbol{o}_{i,t}$, and $Z_{i,t} = 1$, conditioned on the values of the peak boundaries, $B_{i,L,t}$ and $B_{i,R,t}$, and the covariates at time point t, $\boldsymbol{x}_{i,t}$, under the model is then:

$$\begin{aligned} &P(\boldsymbol{O}_{i,t} = \boldsymbol{o}_{i,t}, Z_{i,t} = 1 | B_{i,L,t} = l, B_{i,R,t} = r, \boldsymbol{X}_{i,t} = \boldsymbol{x}_{i,t}) \\ &= P(Z_{i,t} = 1 | B_{i,L,t} = l, B_{i,R,t} = r) \\ &\times \prod_{p=1}^{l-1} \text{NB}(o_{i,t,p}; \mu_{\text{BACKGROUND},i,t,p} = \exp\left[\beta_t + \gamma_t \log \lambda_{i,t,p}\right], \delta_t) \\ &\times \prod_{p=l}^{r} \text{NB}(o_{i,t,p}; \mu_{\text{PEAK},i,t,p} = \exp\left[\alpha_t + \gamma_t \log \lambda_{i,t,p}\right], \delta_t) \\ &\times \prod_{p=r+1}^{N_i} \text{NB}(o_{i,t,p}; \mu_{\text{BACKGROUND},i,t,p} = \exp\left[\beta_t + \gamma_t \log \lambda_{i,t,p}\right], \delta_t) \end{aligned}$$

An important special case of the above formulation when $B_{i,L,t} = B_{i,R,t} + 1$ corresponds to modeling the whole signal at time point t as background, which enables ChromTime to accommodate time points that are all background by modeling them with zero length peaks. For this reason, ChromTime blocks internally have the same number of peak boundaries at all time points even if some time points are predicted as zero length peaks (i.e., all background). Boundaries of zero length peaks are treated by the model in the same way as boundaries of non-zero length peaks.

ChromTime assumes uniform prior probabilities for the left and the right end boundaries at the first time point:

$$P\left(B_{i,L,1} = l\right) = \mathit{Unif}\left(1, N_i + 1\right)$$

and

$$P\left(B_{i,R,1} = r\right) = \mathit{Unif}\left(0, N_i\right)$$

where $\mathit{Unif}(a, b)$ denotes the uniform distribution of integer numbers in the closed interval $[a, b]$.

Let $D_{i,s,t}$ denote the dynamic between time points t and $t + 1$ on boundary side s, where s is one of L (left side) or R (right side). Between any two time points the ChromTime model allows for one of three possible dynamics at both the left and the right end boundaries of a peak: STEADY, EXPAND, or CONTRACT. To capture the change of boundary positions between consecutive time points t and $t + 1$ we define the quantities $J_{i,L,t} = B_{i,L,t} - B_{i,L,t+1}$ and $J_{i,R,t} = B_{i,R,t+1} - B_{i,R,t}$ corresponding to the left and right boundaries, respectively. Positive values of $J_{i,L,t}$ and $J_{i,R,t}$ indicate the number of bases a peak expanded, whereas negative values indicate the number of bases a peak contracted, and a value of 0 indicates that the peak held steady on the left and the right side, respectively. ChromTime models $J_{i,L,t}$ and $J_{i,R,t}$ with different

probability distributions for each of the three dynamics. For STEADY dynamic, ChromTime uses the Kronecker delta function:

$$P\left(J_{i,s,t}|D_{i,s,t}=\mathrm{STEADY}\right)=\begin{cases}1, if\ J_{i,s,t}=0\\ 0, otherwise\end{cases}$$

For expanding and contracting dynamics, ChromTime employs negative binomial distributions to model the number of genomic bins a peak boundary moves relative to the minimal movement of one bin required for peak expansions and contractions:

$$P\left(J_{i,s,t}=j|D_{i,s,t}=\mathrm{EXPAND}\right)=NB\left(j-1;\mu_{\mathrm{EXPAND,t}},\delta_{\mathrm{EXPAND,t}}\right)$$

and

$$P\left(J_{i,s,t}=j|D_{i,s,t}=\mathrm{CONTRACT}\right)=N\mathrm{B}\left(-j-1;\mu_{\mathrm{CONTRACT,t}},\delta_{\mathrm{CONTRACT,t}}\right)$$

Furthermore, each distribution is parametrized with a mean and dispersion parameter depending on the dynamic and the time point, t: $\mu_{\mathrm{EXPAND,t}}$, $\delta_{\mathrm{EXPAND,t}}$ for expansions, and $\mu_{\mathrm{CONTRACT,t}}$, $\delta_{\mathrm{CONTRACT,t}}$ for contractions. Of note, in negative binomial distributions the probabilities for negative integers are defined to be 0. Therefore, the above parametrization enforces that boundary movements of negative or zero length (i.e., contracting or steady, respectively) are impossible for expansions and that boundary movements of positive or zero length (i.e., expanding or steady) are impossible for contractions.

The ChromTime model additionally assumes that there is a prior probability to observe each dynamic between time points t and $t+1$, $P(D_{i,s,t}=d)=\pi_{t,d}$, which is the same at each side (left and right). Users have the option to set a minimum prior probability (MIN_PRIOR) for the dynamics for all time points. This parameter can be used to avoid learning priors too close to zero, which in some cases can occur for more punctate marks where the short length of the peaks can cause the prior to become a dominant influence on the class assignment of the spatial dynamics. By default, MIN_PRIOR = 0 in narrow and broad modes and MIN_PRIOR = 0.05 in punctate mode.

For a time course with T time points we can express for block i the probability of a particular sequence of dynamics and boundary positions on the left side ($\boldsymbol{d}_L$ and $\boldsymbol{b}_L$, respectively) and on the right side ($\boldsymbol{d}_R$ and $\boldsymbol{b}_R$, respectively), and observing foreground counts $\boldsymbol{o}_i$ and $\boldsymbol{Z}_i = \mathbf{1}$ conditioned on the values of the covariates, $\boldsymbol{x}_i$ as:

$$\begin{aligned}&P\left(\boldsymbol{D}_{i,L}=\boldsymbol{d}_L,\boldsymbol{B}_{i,L}=\boldsymbol{b}_L,\boldsymbol{D}_{i,R}=\boldsymbol{d}_R,\boldsymbol{B}_{i,R}=\boldsymbol{b}_R,\boldsymbol{O}_i=\boldsymbol{o}_i,\boldsymbol{Z}_i=\mathbf{1}|\boldsymbol{X}_i=\boldsymbol{x}_i\right)\\&\quad=P\left(B_{i,L,1}=l_1\right)\times P\left(B_{i,R,1}=r_1\right)\\&\quad\times P\left(\boldsymbol{O}_{i,1}=\boldsymbol{o}_{i,1},Z_{i,1}=1|B_{i,L,1}=l_1,B_{i,R,1}=r_1,\boldsymbol{X}_{i,1}=\boldsymbol{x}_{i,1}\right)\\&\quad\times\prod_{t=2}^{T}\left(P\left(\boldsymbol{O}_{i,t}=\boldsymbol{o}_{i,t},Z_{i,t}=1|B_{i,L,t}=l_t,B_{i,R,t}=r_t,\boldsymbol{X}_{i,t}=\boldsymbol{x}_{i,t}\right)\right.\\&\quad\times P\left(J_{i,L,t-1}=l_{t-1}-l_t|D_{i,L,t-1}=d_{L,t-1}\right)\\&\quad\times P\left(D_{i,L,t-1}=d_{L,t-1}\right)\times P\left(J_{i,R,t-1}=r_t-r_{t-1}|D_{i,R,t-1}=d_{R,t-1}\right)\\&\quad\left.\times P\left(D_{i,R,t-1}=d_{R,t-1}\right)\right)\end{aligned}$$

where $\boldsymbol{Z}_i = \mathbf{1}$ is used to denote $Z_{i,t}=1$ for all t, $d_{s,t}$ for $t = 1,\ldots,T\text{-}1$ is the dynamic label for the t^{th} pair of consecutive time points on the left or the right side ($s = L$ or R), respectively. Also $\boldsymbol{b}_L$ and $\boldsymbol{b}_R$ are the vectors of T boundary positions containing l_t and r_t for $t = 1,\ldots,T$, respectively.

The total probability of the signal in a block can be expressed as a sum over all possible sequences of dynamics and peak boundary positions that can generate the block across all time points. Thus, the probability of block i having observations $\boldsymbol{o}_i$ and $\boldsymbol{Z}_i = \mathbf{1}$ given the covariates $\boldsymbol{x}_i$ is:

$$\begin{aligned}&P\left(\boldsymbol{O}_i=\boldsymbol{o}_i,\boldsymbol{Z}_i=\mathbf{1}|\boldsymbol{X}_i=\boldsymbol{x}_i\right)\\&\quad=\sum_{\boldsymbol{d}_L,\boldsymbol{b}_L,\boldsymbol{d}_R,\boldsymbol{b}_R}P\left(\boldsymbol{D}_{i,L}=\boldsymbol{d}_L,\boldsymbol{B}_{i,L}=\boldsymbol{b}_L,\boldsymbol{D}_{i,R}=\boldsymbol{d}_R,\right.\\&\qquad\left.\boldsymbol{B}_{i,R}=\boldsymbol{b}_R,\boldsymbol{O}_i=\boldsymbol{o}_i,\boldsymbol{Z}_i=\mathbf{1}|\boldsymbol{X}_i=\boldsymbol{x}_i\right)\end{aligned}$$

where $\boldsymbol{d}_L$ and $\boldsymbol{d}_R$ each iterate over all possible $3^{T\text{-}1}$ combinations of peak boundary dynamics, and $\boldsymbol{b}_L$ and $\boldsymbol{b}_R$ each iterate over all possible ways to place left and right end boundaries across all time points that are consistent with the requirements that $1 \le B_{i,L,t} \le B_{i,R,t}+1 \le N_i+1$ at each time point.

Let $\boldsymbol{o}$ be the total set of observed read counts in all blocks in the data, $\boldsymbol{x}$ be the set of the corresponding two-dimensional vectors containing the constant term and the logarithm of the expected number of reads at each position and time point for each block, $\boldsymbol{Z} = \mathbf{1}$ denotes all $\boldsymbol{Z_i} = \mathbf{1}$, and M be the total number of blocks. Then, the likelihood of all blocks conditioned on their covariates is:

$$P(\boldsymbol{O}=\boldsymbol{o},\boldsymbol{Z}=\mathbf{1}|\boldsymbol{X}=\boldsymbol{x})=\prod_{i=1}^{M}P\left(\boldsymbol{O}_i=\boldsymbol{o}_i,\boldsymbol{Z}_i=\mathbf{1}|\boldsymbol{X}_i=\boldsymbol{x}_i\right)$$

We note that the above formulation allows ChromTime to model the appearance of a peak, if it occurs after the first time point in the time course, as an expansion from a zero length peak at the previous time point. Similarly, the disappearance of a peak is modeled as a contraction to a zero length peak at the next time point.

Model optimization

The total set of parameters of the model consists of:

1. Prior probabilities of each dynamic d at each time point t: $\pi_{t,d}$.
2. Parameters of the negative binomial distributions that model the PEAK and the BACKGROUND components at each time point: α_t, β_t, γ_t and δ_t.
3. Parameters of the negative binomial distributions that model the boundary movements in EXPAND and CONTRACT dynamics at each time point: $\mu_{\mathrm{EXPAND},t}$, $\delta_{\mathrm{EXPAND},t}$ and $\mu_{\mathrm{CONTRACT},t}$, $\delta_{\mathrm{CONTRACT},t}$, respectively.

The optimal parameter values are attempted to be estimated by Expectation Maximization (EM). In particular, ChromTime attempts to optimize the conditional log-likelihood of the observed counts and $\boldsymbol{Z}_i = \mathbf{1}$ given the covariates (Additional file 2: Supplementary Methods):

$$\sum_{i=1}^{M} \log P(\boldsymbol{O}_i = \boldsymbol{o}_i, \boldsymbol{Z}_i = \mathbf{1} | \boldsymbol{X}_i = \boldsymbol{x}_i)$$

Computing the most likely spatial dynamic and peak boundaries for each block across the whole time course

After the optimal values for all model parameters are estimated from the data, for each block the most likely positions of the peak boundaries at each time point are calculated. This procedure consists of two steps. First, ChromTime determines for each block all time points with significantly low probability of containing a false positive non-zero length peak. Second, conditioned on those time points, ChromTime computes the most likely assignment of the peak boundary variables at each side and each time point (Additional file 2: Supplementary methods).

ChromTime options used in this study

In this work, we applied ChromTime in narrow mode on all data for H3K4me2, H3K4me3, H3K27ac, and H3K(9,14)ac marks. We applied ChromTime in punctate mode on all ATAC-seq and DNase-seq data. No control reads were used for ATAC-seq and DNase-seq. In addition, foreground reads for ATAC-seq were shifted by 5 bp in the direction of alignment (SHIFT = 5), and for DNase-seq no shifting was applied (SHIFT = 0). We applied ChromTime in broad mode on all data for H3K79me2, Pol2, H3K4me1, H3K27me3, H3K9me3, and H3K36me3 marks. All other options were set to their default values.

Timing evaluation

The timing evaluation was conducted on a MacBook Pro laptop with 2.7GHz Intel Core i7 and 16 GB RAM using four CPU cores.

Analyses with external data

The procedures for analyses with external data are described in Additional file 2: Supplementary methods.

Additional files

Additional file 1: Additional figures supporting the main analyses. (PDF 8541 kb)

Additional file 2: Further description of methods and analyses in this study. (PDF 711 kb)

Abbreviations

ATAC-seq: Assay for transposase accessible chromatin coupled with high-throughput DNA sequencing; ChIP-seq: Chromatin immunoprecipitation coupled with high-throughput DNA sequencing; DHS: DNase I hypersensitive site; DNase-seq: DNase I hypersensitivity assay followed by high-throughput DNA sequencing; EM: Expectation maximization; FDR: False discovery rate; HM: Histone mark; TF: Transcription factor; TSS: Transcription start site

Acknowledgements

We are grateful to Constantinos Chronis, Kathrin Plath, and members of the Ernst lab for useful discussions.

Funding

This work was supported by the CIRM Training Grant TG2–01169, the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA Training Program (P.F.); NIH grants R01ES024995, U01HG007912, DP1DA044371, an NSF CAREER Award #1254200 and an Alfred P. Sloan Fellowship (J.E.).

Availability of data and materials

ChromTime software is released under GPL v3.0 and can be downloaded freely at: https://github.com/ernstlab/ChromTime

The version of the code used to perform all analyses in this manuscript is available at: https://doi.org/10.5281/zenodo.1219895 [60].

No new experimental datasets were generated within this study. We used the following publicly available datasets in applying or evaluating ChromTime:

1. ChIP-seq data for histone marks from adipogenesis time course in human and mouse (GEO GSE20752 [18]).
2. ChIP-seq data for histone marks and ATAC-seq data from a blood formation time course in mouse (GEO GSE60103 [19]).
3. ATAC-seq data from a blood formation time course in human. Data from all available healthy donors was pooled for each cell type from the hematopoietic tree (GEO GSE74912 [20]).
4. Aligned reads from DNase-seq experiments from three samples representing human fetal brain development from the Roadmap Epigenomics project (epigenome ids E003, E007, E082 [61–63]).
5. ChIP-seq data for histone marks and Pol2 from a cardiac development time course in mouse (GNomEx database accession number 44R [23, 64]).
6. ChIP-seq data for histone marks and Pol2 from a cardiac development time course in human (GEO GSE35583 [22]).
7. ChIP-seq data for histone marks from an embryogenesis time course in zebrafish (GEO GSE32483 [28]).
8. ChIP-seq data for histone marks from a macrophage differentiation time course in mouse (GEO GSE69101 [21]).
9. ChIP-seq data for histone marks from a neural differentiation time course in human (GEO GSE62193 [12]).

10. ChIP-seq data for histone marks from a stem cell reprogramming time course in human (replicate 1 for all marks and time points and pooled input DNA from all available time points as control, GEO GSE71033 [24]).
11. ChIP-seq data for histone marks and transcription factors, ATAC-seq data, and gene expression data from a stem cell reprogramming time course in mouse (GEO GSE90895 [27]).
12. ChIP-seq data for histone marks and Pol2 from a stem cell reprogramming time course in mouse (GEO GSE67520 [25]).
13. ChIP-seq data for histone marks and GATA3 transcription factor from a T-cell development time course in mouse (GEO GSE31235 [17]).
14. ChIP-seq peaks for OCT4 transcription factor in H1 human embryonic stem cells from the ENCODE project [6, 65].
15. ChIP-seq peaks for NANOG transcription factor in H1 human embryonic stem cells from the ENCODE project [6, 66].
16. ChIP-seq peaks for P300 in H1 human embryonic stem cells from the ENCODE project [6, 67].
17. ChIP-seq peaks for P300 in IMR90 cells were downloaded from ChIP-Atlas [68] at FDR 0.05 [69, 70].
18. ChIP-seq peaks for CEBP in H1 human embryonic stem cells from the ENCODE project [71].
19. ChIP-seq peaks for CEBP in IMR90 cells from the ENCODE project [72].
20. ChIP-seq peaks for Pol2 in H1 human embryonic stem cells from the ENCODE project [73].
21. ChIP-seq peaks for Pol2 in IMR90 cells from the ENCODE project [74].
22. ChIP-seq peaks for Rad21 in H1 human embryonic stem cells from the ENCODE project [75].
23. ChIP-seq peaks for Rad21 in IMR90 cells from the ENCODE project [76].
24. DNase-seq peaks for IMR90 cells from the Roadmap Epigenomics project (epigenome id E017 [77])
25. DNase-seq peaks for H1 human embryonic stem cells from the Roadmap Epigenomics project (epigenome id E003 [78])
26. Gene expression data from the Roadmap Epigenomics project (epigenome ids E003, E007, E082) [79].

Authors' contributions
JE conceived and supervised the project, designed the method, proposed analyses, and wrote the manuscript. PF designed the method, implemented the software, proposed analyses, performed all analyses, processed all data, and wrote the manuscript. Both authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA, USA. [2]Department of Biological Chemistry, University of California, Los Angeles, CA, USA. [3]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA, Los Angeles, CA, USA. [4]Computer Science Department, University of California, Los Angeles, CA, USA. [5]Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA, USA. [6]Molecular Biology Institute, University of California, Los Angeles, CA, USA.

Received: 27 December 2017 Accepted: 17 July 2018
Published online: 10 August 2018

References
1. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008;132:311–22. https://doi.org/10.1016/j.cell.2007.12.014.
2. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10:1213–8. https://doi.org/10.1038/nmeth.2688.
3. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473:43–9. https://doi.org/10.1038/nature09906.
4. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129:823–37. https://doi.org/10.1016/j.cell.2007.05.009.
5. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448:553–60. https://doi.org/10.1038/nature06008.
6. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74. https://doi.org/10.1038/nature11247.
7. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317–30. https://doi.org/10.1038/nature14248.
8. Martens JHA, Stunnenberg HG. BLUEPRINT: mapping human blood cell epigenomes. Haematologica. 2013;98:1487–9. https://doi.org/10.3324/haematol.2013.094243.
9. Lay FD, Triche TJ, Tsai YC, Su S-F, Martin SE, Daneshmand S, et al. Reprogramming of the human intestinal epigenome by surgical tissue transposition. Genome Res. 2014;24:545–53. https://doi.org/10.1101/gr.166439.113.
10. Fiziev P, Akdemir KC, Miller JP, Keung EZ, Samant NS, Sharma S, et al. Systematic epigenomic analysis reveals chromatin states associated with melanoma progression. Cell Rep. 2017;19:875–89. https://doi.org/10.1016/j.celrep.2017.03.078.
11. Mei S, Qin Q, Wu Q, Sun H, Zheng R, Zang C, et al. Cistrome data browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. Nucleic Acids Res. 2017;45:D658–62. https://doi.org/10.1093/nar/gkw983.
12. Ziller MJ, Edri R, Yaffe Y, Donaghey J, Pop R, Mallard W, et al. Dissecting neural differentiation regulatory networks through epigenetic footprinting. Nature. 2014;518:355–9. https://doi.org/10.1038/nature13990.
13. Tsankov AM, Gu H, Akopian V, Ziller MJ, Donaghey J, Amit I, et al. Transcription factor binding dynamics during human ES cell differentiation. Nature. 2015;518:344–9. https://doi.org/10.1038/nature14233.
14. Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2014;518:337–43. https://doi.org/10.1038/nature13835.
15. Gjoneska E, Pfenning AR, Mathys H, Quon G, Kundaje A, Tsai L-H, et al. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. Nature. 2015;518:365–9. https://doi.org/10.1038/nature14252.
16. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson BJ, Xu H, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am J Hum Genet. 2014;95:535–52. https://doi.org/10.1016/j.ajhg.2014.10.004.
17. Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell. 2012;149:467–82. https://doi.org/10.1016/j.cell.2012.01.056.
18. Mikkelsen TS, Xu Z, Zhang X, Wang L, Gimble JM, Lander ES, et al. Comparative epigenomic analysis of murine and human adipogenesis. Cell. 2010;143:156–69. https://doi.org/10.1016/j.cell.2010.09.006.
19. Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretsky I, Jaitin DA, David E, et al. Chromatin state dynamics during blood formation. Science. 2014;345:943–9. https://doi.org/10.1126/science.1256271.
20. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, et al. Lineage-specific and single-cell chromatin accessibility charts human

hematopoiesis and leukemia evolution. Nat Genet. 2016;48:1193–203. https://doi.org/10.1038/ng.3646.
21. Goode DK, Obier N, Vijayabaskar MS, Lie-A-Ling M, Lilly AJ, Hannah R, et al. Dynamic gene regulatory networks drive hematopoietic specification and differentiation. Dev Cell. 2016; https://doi.org/10.1016/j.devcel.2016.01.024.
22. Paige SL, Thomas S, Stoick-Cooper CL, Wang H, Maves L, Sandstrom R, et al. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell. 2012;151:221–32. https://doi.org/10.1016/j.cell.2012.08.027.
23. Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell. 2012;151:206–20. https://doi.org/10.1016/j.cell.2012.07.035.
24. Cacchiarelli D, Trapnell C, Ziller MJ, Soumillon M, Cesana M, Karnik R, et al. Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. Cell. 2015;162:412–24. https://doi.org/10.1016/j.cell.2015.06.016.
25. Chen J, Chen X, Li M, Liu X, Gao Y, Kou X, et al. Hierarchical Oct4 binding in concert with primed epigenetic rearrangements during somatic cell reprogramming. Cell Rep. 2016; https://doi.org/10.1016/j.celrep.2016.01.013.
26. Koche RP, Smith ZD, Adli M, Gu H, Ku M, Gnirke A, et al. Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell. 2011;8:96–105. https://doi.org/10.1016/j.stem.2010.12.001.
27. Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, et al. Cooperative binding of transcription factors orchestrates reprogramming. Cell. 2017;168:442–459.e20. https://doi.org/10.1016/j.cell.2016.12.016.
28. Bogdanovic O, Fernandez-Miñán A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Kruysbergen I, et al. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res. 2012;22:2043–53. https://doi.org/10.1101/gr.134833.111.
29. Nègre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, et al. A cis-regulatory map of the Drosophila genome. Nature. 2011;471:527–31. https://doi.org/10.1038/nature09990.
30. Yu P, Xiao S, Xin X, Song C-X, Huang W, McDee D, et al. Spatiotemporal clustering of epigenome reveals rules of dynamic gene regulation. Genome Res. 2013;23:352–64. doi:https://doi.org/10.1101/gr.144949.112.
31. Arnold P, Schöler A, Pachkov M, Balwierz P, Jørgensen H, Stadler MB, et al. Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. Genome Res. 2013;23:60–73. https://doi.org/10.1101/gr.142661.112.
32. Koike N, Yoo S-H, Huang H-C, Kumar V, Lee C, Kim T-K, et al. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. Science. 2012;338:349–54. https://doi.org/10.1126/science.1226339.
33. Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, et al. Latent enhancers activated by stimulation in differentiated cells. Cell. 2013;152:157–71. https://doi.org/10.1016/j.cell.2012.12.018.
34. Reilly SK, Yin J, Ayoub AE, Emera D, Leng J, Cotney J, et al. Evolutionary changes in promoter and enhancer activity during human corticogenesis. Science. 2015;347:1155–9. https://doi.org/10.1126/science.1260943.
35. Weiner A, Hsieh T-HS, Appleboim A, Chen HV, Rahat A, Amit I, et al. High-resolution chromatin dynamics during a yeast stress response. Mol Cell. 2015;58:371–86. https://doi.org/10.1016/j.molcel.2015.02.002.
36. Cotney J, Leng J, Yin J, Reilly SK, DeMare LE, Emera D, et al. The evolution of lineage-specific regulatory activities in the human embryonic limb. Cell. 2013;154:185–96. https://doi.org/10.1016/j.cell.2013.05.056.
37. Zhu J, Wang J, Chen X, Tsompana M, Gaile D, Buck M, et al. A time-series analysis of altered histone H3 acetylation and gene expression during the course of MMAIII-induced malignant transformation of urinary bladder cells. Carcinogenesis. 2017;38:378–90. https://doi.org/10.1093/carcin/bgx011.
38. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137. https://doi.org/10.1186/gb-2008-9-9-r137.
39. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576–89. https://doi.org/10.1016/j.molcel.2010.05.004.
40. Xu S, Grullon S, Ge K, Peng W. Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. Methods Mol Biol. 2014;1150:97–111. https://doi.org/10.1007/978-1-4939-0512-6_5.
41. Xing H, Mo Y, Liao W, Zhang MQ, Ren B, Robert F, et al. Genome-wide localization of protein-DNA binding and histone modification by a Bayesian change-point method with ChIP-seq data. PLoS Comput Biol. 2012;8:e1002613. https://doi.org/10.1371/journal.pcbi.1002613.
42. Harmanci A, Rozowsky J, Gerstein M. MUSIC: identification of enriched regions in ChIP-Seq experiments using a mappability-corrected multiscale signal processing framework. Genome Biol. 2014;15:474. https://doi.org/10.1186/s13059-014-0474-3.
43. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 2012;9:215–6. https://doi.org/10.1038/nmeth.1906.
44. Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. Nat Methods. 2012;9:473–6. https://doi.org/10.1038/nmeth.1937.
45. Allhoff M, Seré K, Chauvistré H, Lin Q, Zenke M, Costa IG. Detecting differential peaks in ChIP-seq signals with ODIN. Bioinformatics. 2014;30:3467–75. https://doi.org/10.1093/bioinformatics/btu722.
46. Xu H, Wei C-L, Lin F, Sung W-K. An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data. Bioinformatics. 2008;24:2344–9. https://doi.org/10.1093/bioinformatics/btn402.
47. Biesinger J, Wang Y, Xie X. Discovering and mapping chromatin states using a tree hidden Markov model. BMC Bioinf. 2013;14 Suppl 5:S4. https://doi.org/10.1186/1471-2105-14-S5-S4.
48. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155:934–47. https://doi.org/10.1016/j.cell.2013.09.053.
49. Parker SCJ, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci U S A. 2013;110:17921–6. https://doi.org/10.1073/pnas.1317023110.
50. Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, et al. H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell. 2014;158:673–88. https://doi.org/10.1016/j.cell.2014.06.027.
51. Chen K, Chen Z, Wu D, Zhang L, Lin X, Su J, et al. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. Nat Genet. 2015;47:1149–57. https://doi.org/10.1038/ng.3385.
52. Dincer A, Gavin DP, Xu K, Zhang B, Dudley JT, Schadt EE, et al. Deciphering H3K4me3 broad domains associated with gene-regulatory networks and conserved epigenomic landscapes in the human brain. Transl Psychiatry. 2015;5:e679. https://doi.org/10.1038/tp.2015.169.
53. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell. 2010;6:479–91. https://doi.org/10.1016/j.stem.2010.03.018.
54. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. 2008;40:897–903. https://doi.org/10.1038/ng.154.
55. Ng HH, Robert F, Young RA, Struhl K. Targeted recruitment of Set1 histone methylase by elongating pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell. 2003;11:709–19.
56. Smith E, Lin C, Shilatifard A. The super elongation complex (SEC) and MLL in development and disease. Genes Dev. 2011;25:661–72. https://doi.org/10.1101/gad.2015411.
57. Lacadie SA, Ibrahim MM, Gokhale SA, Ohler U. Divergent transcription and epigenetic directionality of human promoters. FEBS J. 2016; https://doi.org/10.1111/febs.13747.
58. Cameron AC, Trivedi PK. Regression analysis of count data. Second edition. Cambridge: Cambridge University Press; 2013.
59. Rashid NU, Giresi PG, Ibrahim JG, Sun W, Lieb JD. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biol. 2011;12:R67. https://doi.org/10.1186/gb-2011-12-7-r67.
60. Fiziev: ernstlab/ChromTime: ChromTime v1.0.0 [Code] Zenodo. 2018. https://doi.org/10.5281/zenodo.1219895.
61. Roadmap Epigenomics Consortium. DNase-seq data for H1 human embryonic stem cells. http://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/E003-DNase.tagAlign.gz. Accessed 24 Jun 2018.
62. Roadmap Epigenomics Consortium. DNase-seq data for H1-derived neuronal progenitors. http://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/E007-DNase.tagAlign.gz. Accessed 24 Jun 2018.

63. Roadmap Epigenomics Consortium. DNase-seq data for fetal brain tissue. http://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/E082-DNase.tagAlign.gz. Accessed 24 Jun 2018.
64. Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, et al. ChIP-seq data for histone marks. https://hci-bio-app.hci.utah.edu/gnomex/. Accessed 11 Sep 2017.
65. ENCODE Project Consortium. ChIP-seq peaks for OCT4 transcription factor in H1 human embryonic stem cells. https://www.encodeproject.org/files/ENCFF002CJF/@@download/ENCFF002CJF.bed.gz. Accessed 24 Jun 2018.
66. ENCODE Project Consortium. ChIP-seq peaks for NANOG transcription factor in H1 human embryonic stem cells. https://www.encodeproject.org/files/ENCFF002CJA/@@download/ENCFF002CJA.bed.gz. Accessed 24 Jun 2018.
67. ENCODE Project Consortium. ChIP-seq peaks for P300 in H1 human embryonic stem cells. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/wgEncodeBroadHistoneH1hescP300kat3bPk.broadPeak.gz. Accessed 24 Jun 2018.
68. Oki S, Ohta T, Shioi G, Hatanaka H, Ogasawara O, Okuda Y, et al. Integrative analysis of transcription factor occupancy at enhancers and disease risk loci in noncoding genomic regions. bioRxiv. 2018:262899. https://doi.org/10.1101/262899.
69. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature. 2013; 503:290–4. https://doi.org/10.1038/nature12644.
70. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. ChIP-seq peaks for P300 in IMR90 cells at FDR 0.05. http://chip-atlas.org/view?id=SRX212184. Accessed 24 Jun 2018.
71. ENCODE Project Consortium. ChIP-seq peaks for CEBP in H1 human embryonic stem cells. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeAwgTfbsUniform/wgEncodeAwgTfbsSydhH1hescCebpbIggrabUniPk.narrowPeak.gz. Accessed 24 Jun 2018.
72. ENCODE Project Consortium. ChIP-seq peaks for CEBP in IMR90 cells. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeAwgTfbsUniform/wgEncodeAwgTfbsSydhImr90CebpbIggrabUniPk.narrowPeak.gz. Accessed 24 Jun 2018.
73. ENCODE Project Consortium. ChIP-seq peaks for Pol2 in H1 human embryonic stem cells. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeAwgTfbsUniform/wgEncodeAwgTfbsUtaH1hescPol2UniPk.narrowPeak.gz. Accessed 24 Jun 2018.
74. ENCODE Project Consortium. ChIP-seq peaks for Pol2 in IMR90 cells. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeAwgTfbsUniform/wgEncodeAwgTfbsSydhImr90Pol2IggrabUniPk.narrowPeak.gz. Accessed 24 Jun 2018.
75. ENCODE Project Consortium. ChIP-seq peaks for Rad21 in H1 human embryonic stem cells. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeAwgTfbsUniform/wgEncodeAwgTfbsSydhH1hescRad21IggrabUniPk.narrowPeak.gz. Accessed 24 Jun 2018.
76. ENCODE Project Consortium. ChIP-seq peaks for Rad21 in IMR90 cells. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeAwgTfbsUniform/wgEncodeAwgTfbsSydhImr90Rad21IggrabUniPk.narrowPeak.gz. Accessed 24 Jun 2018.
77. Roadmap Epigenomics Consortium. DNase-seq peaks for IMR90 cells. http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/narrowPeak/E017-DNase.macs2.narrowPeak.gz. Accessed 24 Jun 2018.
78. Roadmap Epigenomics Consortium. DNase-seq peaks for H1 human embryonic stem cells. http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/narrowPeak/E003-DNase.macs2.narrowPeak.gz. Accessed 24 Jun 2018.
79. Roadmap Epigenomics Consortium. Gene expression data for protein coding genes in 57 samples. http://egg2.wustl.edu/roadmap/data/byDataType/rna/expression/57epigenomes.RPKM.pc.gz. Accessed 24 Jun 2018.
80. Bar-Joseph Z, Gifford DK, Jaakkola TS. Fast optimal leaf ordering for hierarchical clustering. Bioinformatics. 2001;17(SUPPL. 1):S22–9. https://doi.org/10.1093/bioinformatics/17.suppl_1.S22.